Illinois State Water Survey Division

ATMOSPHERIC CHEMISTRY SECTION

AT THE

UNIVERSITYOF ILLINOIS

SWS Contract Report 420

ATMOSPHERIC DEPOSITION: EVALUATION OF AREAL MEAN ESTIMATION METHODS

by Chin-Fei Hsu

Final Report

Grant INT 14-08-0001-G-964

U.S. Geological Survey

Principal Investigators: Chin-Fei Hsu and Donald F. Gatz

Champaign, Illinois

May 1987

Illinois Department of Energy and Natural Resources

CONTENTS

ATMOSPHERIC DEPOSITION: EVALUATION OF AREAL MEAN ESTIMATION METHODS

by Chin-Fei Hsu

INTRODUCTION

This report summarizes the efforts and results of a study performed by the U.S. Geological Survey under grant INT 14-08-0001-G-964, entitled: "Atmospheric Deposition: Evaluation of Areal Mean Estimation Methods."

Statement of the Problem

The National Atmospheric Deposition Program (NADP) was organized in part to measure atmospheric deposition and its spatial and temporal trends on regional and national scales (Galloway and Cowling, 1978). Spatial trends can be inferred by comparing mean depositions between regions for the same time period. Temporal trends can be assessed by comparing mean deposition over given regions between years. However, atmospheric deposition cannot be measured directly, but must be estimated from sample measurements at a few discrete collectors, and for time periods of a few years or less.

The area of a collector is a very small fraction of the area it is presumed to represent. Estimations of atmospheric deposition therefore have uncertainties associated with them. We need to know how to express these uncertainties properly in order to attain the NADP and National Trends Network (NTN) goals of measuring temporal and spatial variations of atmospheric deposition.

OBJECTIVES AND PROCEDURES

The purpose of this study was to estimate quantitative error limits on areal mean deposition. The research had two primary objectives:

1. To study methods of estimating the uncertainty of areal-temporal mean deposition

2. To apply the methods to the current NADP data

To carry out these objectives, we studied several statistical methods of estimating the uncertainty of areal-temporal mean deposition. These methods were temporal-spatial correlation analysis, kriging, and time series analysis. In the first half of the study period, we studied the statistical methods, and in the second half, we applied these methods to the NADP data.

METHODS OF ESTIMATING AREAL MEAN

Suppose S is a region, and [0,T] is a time period for which we want to estimate the areal-temporal mean characteristic of a variable, say deposition of an ion,

$$Z_0 = \frac{1}{TA(S)} \int_0^T \int_S dZ(x,y,t) \tag{1}$$

where A(S) is the area of region S, and Z(x,y,t) is the (stochastic) deposition process of the ion at the point (x,y) at time t. To estimate ZQ, a network of a finite number (N) of collectors is used, and observations z(x,y,t) are taken at collector site (x,y) at discrete times, t = 0,1...,T. If we presume that temporal variability is of no interest over the area of study, we estimate the areal mean by

$$Z_1 = \frac{1}{A(S)} \int_S dZ(x,y) \tag{2}$$

where Z(x,y) is a spatial process of the "total" deposition of an ion at the point (x,y). In order to derive the variance for an estimator of ZQ, certain information about the covariance structure of the process Z is required. Some of these conditions and the associated terminology are described in a companion report reprinted as the appendix to this report (Hsu and Guerrero, 1986), in which we describe the following terms: (spatially) weakly stationary process, generalized covariance of order k, (temporally) weakly stationary process, intrinsic hypothesis, semi-variogram, permissible covariance, generalized increment of order k, and intrinsic random function of order k.

Correlation Approach

A model for representing the temporal-spatial covariance function of a surface rainfall process (i.e., rainfall measured at the earth's surface) was proposed by Rodriguez-Iturbe and Mejia (1974). Gatz and Naiman (1980) used this approach to investigate the spatial variability of a number of ions in rainwater. The model is based on the assumption that the stochastic process is weakly stationary in time and space, and the covariance structure of the process is separable into a temporal component and a spatial component:

$$Cov[Z(x,y,t),Z(u,v,s)] = \sigma^2 r(d) r^*(t-s) \tag{3}$$

where σ^2 is the (constant) point variance of Z(x,y,t), d is the distance between (x,y) and (u,v), r(d) is the spatial component of the covariance

function, and r*(t – s) is the temporal component. r*(t – s) has been approximated by a simple Markovian form:

$$r^*(t-s)=\rho^{|t-s|} \tag{4}$$

where p denotes the first-order autocorrelation coefficient of the process. In addition, assuming the process to be isotropic, r(d) can be taken to be either a negative exponential function or a first-order modified Bessel function. (Bras and Rodriguez-Iturbe (1976) provide a more extensive discussion of the form of r(d). Rodriguez-Iturbe and Mejia (1974) present a geometrical method to fit r(d).)

To estimate the areal mean (1), Rodriguez-Iturbe and Mejia (1974) used a simple arithmetic mean and expressed the mean square error (MSE) of this estimator as

$$MSE=E(Z_0-Z_0)^2=\sigma^2F_1(T)F_2(N) \tag{5}$$

where σ^2 is the point variance, F_1 is the variance reduction factor due to temporal sampling, and F_2 is the variance reduction factor due to spatial sampling.

Relevance to the NADP Data. The use of a simple arithmetic mean is based on the implicit assumption that the collectors (stations) are uniformly distributed over the area of interest. Rodriguez-Iturbe and Mejia (1974) and Lenton and Rodriguez-Iturbe (1977) used a weighted average and derived a set of optimum weights which minimizes MSE. The estimate can be further improved by first weighing Z(x,y,t) by area size. Bras and Colon (1978) extended the estimation procedure to allow for the covariance function to be non-separable by adopting a stationary (multivariate) autoregressive model for the underlying process.

One implication of the assumption of spatial stationarity is that the process is relatively homogeneous over the area of interest. In order to satisfy this, the area cannot be too large. A division of the national NADP network into regions of homogeneous variances and covariances was carried out by using the pointwise standard deviation (Hsu and Gatz, 1985). The assumption of isotropicity was used in our attempt to analyze the NADP data.

Kriging Approach

The kriging technique is an extension of the "generalized least squares" technique. General discussion on kriging can be found in Matheron (1971), David (1976), Delhomme (1978), Kafritsas and Bras (1981), and Chirlin and Wood (1982). Recently, Finkelstein and Seilkop (1981) and Finkelstein (1984) applied kriging to a study of acid precipitation, using NADP data. Kriging is used to find, from a "single realization" of Z, the relationship between the

spatial correlation and the precision of interpolation and/or to estimate the mean of a regionalized variable.

To make statistical inference feasible, certain assumptions about Z must be made. The model is based on the assumption that Z is *ergodic,* i.e., the first two moments of Z tend asymptotically to certain constants. The variogram y(d) can also be allowed to be anisotropic, though the procedure then becomes complicated. In addition, sometimes Z is assumed to satisfy spatial (weak) stationarity or intrinsic hypothesis, or to allow for a drift. (See the appendix for descriptions of these terms.) The technique which "differences" the process in order to achieve zero mean (spatial) stationarity is sometimes referred to as "universal kriging" (Delfiner, 1976; Chirlin and Wood, 1982).

To estimate the areal mean (2) of Z over an area S, a weighted average of N point observations is used. The estimator is derived so that it is unbiased and minimizes the MSE. Solving a "kriging system" of equations can show that such an estimator exists. The estimated weights which satisfy the above conditions are called *kriging values,* and the minimum variance the *kriging variance.* If the estimator is an "increment of order k" (Delhomme, 1978), its variance can be shown to be

$$\sum\lambda_i\lambda_j K(d(i,j)) \qquad (6)$$

where d(i,j) is the distance between the i-th and the j-th points, and K( ) is the generalized covariance.

Relevance to the WADP Data. The kriging technique requires less stringent assumptions than the correlation approach or time series approach. The technique allows for spatial non-stationarity and anisotropic covariance structure and is appropriate in estimating long-term areal mean deposition. Finkelstein and Seilkop (1981) preferred kriging over the correlation approach or the time series approach. In order to use the kriging technique, one has to disregard temporal variability, an assumption which might require a fairly long record (i.e., many years) of deposition observations to satisfy.

During the fitting stages, the variogram needs to be calculated (see Chua and Bras, 1980, for details), and the degree of drift as well as the generalized covariance also need to be determined (see Volpi and Gambolati, 1978; and Chua and Bras, 1980). Delhomme (1978) presented a jackknife procedure for checking the validity of the fitted model.

Time Series Approach

The technique of time series analysis has long been applied to the hydrological and atmospheric sciences. Ord and Rees (1979) reviewed the development of spatial time series and discussed the estimation of areal mean.

Given a time series observed at a collector site, significant periodicities can be extracted from the series after it is detrended (or pre-

whitened). The residuals are defined as the differences between the observed value and the fitted value. Most test statistics for periodicity assume a normal distribution for residuals. A goodness-of-fit test for normality is thus required. After the appropriate time series analysis for each site, an estimate of areal mean can be obtained by a method of averaging, and an estimate of the variance can be derived. The use of this approach is reported by Neill and Hsu (1981) and Zinsmeister and Redman (1980).

A multiple time series method can also be used. The interrelation between different sites is useful in reducing the magnitude of residuals and thus improving precision of the estimated areal mean. Rhenals-Figueredo et al. (1974) used a bi-dimensional spectral analysis to model point and areal rainfall, and studied the correlation function of the residuals. Matern (1960) derived formulas for the estimator of an areal mean, its variance, and the corresponding spectral density function for a stationary process, for both isotropic and anisotropic models.

During the course of the study, a spatial time series approach called STARMA was found to be a promising approach for our study. The method was developed by S. J. Deutsch and his students at the Georgia Institute of Technology (Pfeifer, 1979; Pfeifer and Deutsch, 1980a, 1980b, 1980c, 1981). Intensive iterative procedures are needed in using this method, and thus heavy computer usage is necessary.

Considerable effort was given to studying the STARMA method and discussing it with C. Wang, Department of Mechanical and Industrial Engineering at the University of Illinois, who has been a student of Professor Deutsch. A joint effort was proposed in pursuing the use of STARMA in analyzing NADP data. Unfortunately, subsequent events prevented this from taking place. First, the software that performs the STARMA computation could not be used because of "commercial licensing" complications (we could not use the software without the consulting services of Professor Deutsch); second, the STARMA software usage required more involvement of C. Wang than first envisioned, so the analysis could not be completed in the time frame of this research; and lastly, the transfer of C. Wang to Virginia Polytechnic Institute and State University made the joint effort difficult to pursue.

Thus we were able to use the STARMA approach for only a few trials of preliminary analysis using NADP data. Since STARMA is able to take into consideration the relative locations of sites, its results should be better than those of the classical multiple time series method. It is therefore a worthwhile approach to continue to pursue for NADP analysis.

The difficulties associated with the time series approach are model identification and verification of the assumption of (temporal) stationarity.

ANALYSIS OF THE NADP DATA

The correlation approach, kriging approach, and time series approach were used to analyze the NADP data 1) to obtain estimates of various regional areal mean depositions, and 2) to derive estimates of the variance (i.e., uncertainty) of these estimates.

The NADP data used in the analyses were for the period from the summer of 1978 until the end of 1983. The data consisted of weekly (mostly Tuesday-to-Tuesday) values of ion concentrations, conductivity, pH, and rain volume. The data were from 129 collector sites, of which 39 sites were excluded from analysis because they had fewer than 52 weeks of observations. Questionable samples as recorded in the field and laboratory notes were not used, except for values from irregular sampling periods, which were converted to 7-day values where possible. No missing values were estimated.

Exploratory analyses were applied to the data to check the validity of the assumptions of each approach. Division of the NADP network into regions of homogeneous variance was carried out by looking at standard deviations combined with geographical considerations.

The results of the NADP analyses have been presented at two conferences. Preliminary results were presented at the Ninth Conference of Probability and Statistics on Atmospheric Sciences. A copy of this proceeding paper has been submitted to the U.S. Geological Survey. Following this, a talk about the more recent analyses was given at the Muskoka '85 Conference (International Symposium on Acidic Precipitation) in Canada. This talk was subsequently summarized in a paper published in the journal Water. Air and Soil Pollution in 1986. Results of the analyses are summarized as follows:

The continental U.S.A. is divided into three regions (east, central, and west) according to whether the standard deviation of pH is larger than 0.6 or not. The east and west regions have standard deviations smaller than 0.6 and the central region has a standard deviation larger than 0.6. The overall mean pH for the continental U.S. is 4.95. The arithmetic means of conductivity, NO3, and SO4 have a noticeable east-west gradient with the largest values occurring in the east. The reverse is true for C1, pH, and PO_4. For Ca, K, Mg, Na, and NH4, the peak values occur in the central region.

In general, the kriging weighted means are slightly larger than the arithmetic means except for NH4, NO_3, and PO_4. The kriging square root MSE is smaller than the standard deviations of K, Mg, and pH in all regions but is larger than that of SO4. Similarly, the weighted mean calculated from the correlation approach is generally smaller than the arithmetic means. Almost all of the square root MSEs from the correlation approach are smaller than the standard deviations. In all, the weighted means calculated by the correlation approach are smaller than those calculated by kriging. The MSEs from the correlation approach are smaller than those from kriging for K, NH4, NO_3, pH, and volume for all regions; while mixed results occur for other ions.

DISCUSSION

Baker et al. (1979) estimated areal-averaged aerosol concentrations of industrial sulfates and windblown dust by assuming a Markovian rainfall process and deriving the mean square errors of the estimates. Van Egmond and Onderdelinden (1981) compared three interpolation techniques (optimum interpolation, eigenvector interpolation, and distance-density interpolation) and found their differences to be small. They concluded that sampling error and measurement error were of overriding importance in the efficiency of reconstruction of pollution concentration.

Bras and Colon (1978) used a Kalman-Bucey filter to estimate a time-averaged areal mean of precipitation and derived its MSE for a given network in terms of the number of sites, their locations, and a given number of time intervals. Karol and Myatch (1972) used an optimal interpolation scheme (namely, Gandin's objective analysis) to estimate areal mean wet deposition, and compared the MSEs of the estimates for several network configurations and several separation distances between the collector sites.

ACKNOWLEDGEMENTS

We thank Van Bowersox for assistance in many ways in obtaining the NADP data from the University of Colorado and in helping us interpret the data. We also thank Randy Stahlhut for helping us process the data. We appreciate very much the help of Jim Angel in carrying out many programming jobs. Finally, we thank Jose Guerrero, a graduate student, who spent many laborious hours doing a literature search and studying the theory of spatial process.

REFERENCES

Baker, M. B. , H. Harrison, J. Vinelli, and K. B. Erickson, 1979: Simple stochastic models for the sources and sinks of two aerosol types. Tellus. 31:39-51.

Bras, R. L. and I. Rodriguez-Iturbe, 1976: Evaluation of mean square error involved in approximating the areal average of a rainfall event by a discrete summation. Water Resour. Res., 12: (2), 181-184.

Bras, R. L. and R. Colon, 1978: Time-averaged areal mean of precipitation: estimation and network design. Water Resour. Res.. 14:(5), 878-888.

Chirlin, G. R. and E. F. Wood, 1982: On the relationship between kriging and state estimation. Water Resour. Res.. 18: (2), 432-438.

Chua, S. H. and R. L. Bras, 1980: Estimation of stationary and non-stationary random fields: kriging in the analysis of orographic precipitation. Report 255, Dept. Civil Eng., MIT, 171 pp.

David, M. , 1976: The practice of kriging. In Advanced Geostatistics in the Mining Industry, M. Guarascio, M. David, and C. Huijbregts [Eds.], D. Reidel Publ. Co., Boston, 31-48.

Delfiner, P., 1976: Linear estimation of non stationary spatial phenomena. In Advanced Geostatistics in the Mining Industry, M. Guarascio, M. David, and C. Huijbregts [Eds.], D. Reidel Publ. Co., Boston, 49-68.

Delhomme, J. P., 1978: Kriging in the hydrosciences. Adv. Water Res., 1:(5), 251-266.

Finkelstein, P. L. and S. K. Seilkop, 1981: Interpolation error and the spatial variability of acid precipitation. Seventh Conf. on Prob. and Statist. in Atmos. Sci., Amer. Meteor. Soc, 206-212.

Finkelstein, P. L., 1984: Interpolation error and the spatial variability of acid precipitation. J. Clim. and Appl. Meteor., 23. 52.

Galloway, J. N. and E. B. Cowling, 1978: The effects of precipitation over aquatic and terrestrial ecosystems - a proposed precipitation chemistry network. J. Air Pollut. Control Assoc. 28. 229-235.

Gatz, D. F. and D. Q. Naiman, 1980: Spatial variability of rain water impurities in mesoscale events. Symp., Interm. Range Atmos. Transp. Proc. and Tech. Assess., 1-3 Oct., Gatlinburg, TN, 6 pp.

Hsu, C.-F. and D. F. Gatz, 1985: Uncertainties of areal means estimation: applications to rain chemistry data (preliminary results). Ninth Conf. on Prob. and Statist. in Atmos. Sci., Amer. Meteor. Soc, 6 pp.

Hsu, C.-F. and J. Guerrero, 1986: Some aspects of kriging, with application to areal rainfall estimation. Illinois State Water Survey, 13 pp. [see appendix].

Kafritsas, J. and R. L. Bras, 1981: The practice of kriging. Report 263, Dept. Civil Eng., MIT, 107 pp.

Karol, I. L. and L. T. Myatch, 1972: Contribution to the planning of the station network for measuring the chemical composition of precipitation. Tellus. 26: (5), 421-427.

Lenton, R. L. and I. Rodriguez-Iturbe, 1977: Rainfall network systems analysis: the optimal estimation of total areal storm depth. Water Resour. Res., 13:(5), 825-836.

Matern, B. , 1960: Spatial variation. Statens Skogsfors. Inst., 49:(5), 144 PP.

Matheron, G., 1971: The theory of regionalized variables and its applications. Cahiers du Centre de Morphologie Math., 5, Exole des Mines, Fontainebleau, France, 211 pp.

Neill, J. C. and C.-F. Hsu, 1981: Using non-integer spectral analysis in discerning spatially coherent rainfall periodicities. Time Series Analysis, 0. D. Anderson and M. R. Perryma [Eds.], North-Holland Publ. Co.

Ord, K. and M. Rees, 1979: Spatial processes: recent developments with applications to hydrology. Math. Hydrol. and Water Res., F. H. Lloyd, T. O'Donnell, and J. C. Wilkinson [Eds.], Academic Press, 95-118.

Pfeifer, P. E. , 1979: Spatial-dynamic modeling. Ph.D. dissertation, Georgia Institute of Technology, Atlanta, GA.

Pfeifer, P. E. and S. J. Deutsch, 1980a: A STARMA modeling building procedure with application to description and regional forecasting. Trans. Inst. British Geographers, 5.

Pfeifer, P. E. and S. J. Deutsch, 1980b: A three-stage iterative procedure for space-time modeling. Technometries. 22. 35-47.

Pfeifer, P. E. and S. J. Deutsch, 1980c: Identification and interpretation of first-order space-time ARMA models. Technometrics. 22. 397-408.

Pfeifer, P. E. and S. J. Deutsch, 1981: Variance of the sample space-time auto-correlation functions. J. Roy. Stastit. Soc., B, 43:1. 28-33.

Rhenals-Figueredo, A. E. , I. Rodriguez-Iturbe, and J. C. Schaake, Jr., 1974: Bidimensional spectral analysis of rainfall events. Report 193, Dept. Civil Eng., MIT, 214 pp.

Rodriguez-Iturbe, I. and J. M. Mejia, 1974: The design of rainfall networks in time and space. Water Resour. Res.. 10:(4), 713-728.

Van Egmond, N. D. and D. Onderdelinden, 1981: Objective analysis of air pollution monitoring network data; spatial interpolation and network density. Atmos. Environ., 15: (6), 1035-1046.

Volpi, G. and G. Gambolati, 1978: On the use of a main trend for the kriging technique in hydrology. Advances in Water Resources. 1, 345-349.

Zinsmeister, A. R. and T. C. Redman, 1980: A time series analysis of aerosol composition measurements. Atmos. Environ., 14. 205-215.

PROJECT REPORTS AND PUBLICATIONS

Hsu, Chin-Fei and D. F. Gatz, 1985: Uncertainties of areal means estimation: applications to rain chemistry data (preliminary results). Ninth Conf. on Prob. and Statist. in Atmos. Sci., Amer. Meteor. Soc., 6 pp.

Hsu, Chin-Fei, 1985: Uncertainties in estimating areal means: with applications to NADP/NTN data (preliminary). Paper presented at Muskoka '85, International Symposium on Acidic Precipitation, Ontario, Canada.

Hsu, C.-F. and J. Guerrero, 1986: Some aspects of kriging, with application to areal rainfall estimation. Illinois State Water Survey, 13 pp. [see appendix].

Hsu, Chin-Fei, 1986: Uncertainties in estimating areal means: with applications to NADP/NTN data. Water. Air and Soil Pollut., 30, 295-302.

APPENDIX:

SOME ASPECTS OF KRIGING, WITH APPLICATION TO AREAL RAINFALL ESTIMATION

C. F. Hsu and J. Guerrero

Illinois State Water Survey
Champaign, I1 61820

INTRODUCTION

A number of procedures which estimate areal characteristics of certain variables have appeared in the literature. Some of these which dealt with areal rainfall were reviewed by Hall and Barclay (1975). These procedures have an appealing feature in that the variance associated with the estimators can be derived. Therefore by comparing the magnitude of the variances it is possible to select from these methods an estimator with the smallest variance. Some of these procedures utilizes temporal-spatial correlation, some uses time series approach, while others use the 'kriging' approach. The underlying assumptions made in each approach are different and must be verified against a specific kind of data before a method can be used. In this paper, one of these methods, kriging, is studied. The assumptions associated with it are discussed, and examples of its applications to estimate areal rainfall are presented.

Def: A random (or stochastic) process $Z(t)$ is a family of random variables indexed by the symbol t, where t belongs to some given index set T. If t takes a continuous range of real values, $Z(t)$ is said to be a continuous parameter process. If t takes an enumerable set of values, typically $t=...,-1,0,1,...$, then $Z(t)$ is said to be a discrete parameter process (or time series).

A process is called multidimensional when the parameter t has several components $t_1, t_2, \bullet\bullet\bullet, t_p$, say. In this case the parameter t is vector-valued and of course can no longer represent 'time' only, although one of its components could still represent a time parameter. The distinction between a multivariate process and multidimensional process is that in a multivariate process Z is a vector and t is a scalar, whereas in a multidimensional process Z is a scalar and t is a vector. In the following, only multidimensional process will be considered, with emphasis on 2-dimensional process.

One can consider a continuous process in which the variate has a value at every point of the 'domain of interest', or a discrete process in which the variate is observed at a discrete set of points of the hyperplane.

One can also define continuity or discreteness with respect to the random variable under study.

Suppose S is a region and [0,T] is a time period for which we want to estimate the areal-temporal mean characteristic of a variable, for example, rainfall,

$$Z_0 = (1/TA(S)) \int_0^T \int_S dZ(x,y,t), \tag{1}$$

where A(S) is the area of a region S in R^2 (2-dimensional real space) and Z(x,y,t) is the (stochastic) process of the variable at the point (x,y) at time t. In reality, to estimate Zg, a network of a finite number (N) of collectors is used, and observations are taken at each collector site at discrete time, t = 0,1,...,T.

Or if we may suppose that the temporal variability does not matter over the area of interest, we may then simply estimate the areal mean characteristic

$$Z_1 = (1/A(S)) \int_S dZ(x,y), \quad S \text{ in } R^2 \tag{2}$$

where Z(x,y) is a spatial process of 'total' rain at point (x,y) over the entire time period [0,T]. One such example is the estimation of mean areal total rain due to a storm. To estimate Z_1, the same network of collectors as above can be used. However, the sum of observed values over [0,T] are used instead of individual values at discrete time points. In applications such as ore exploration, the value taken at each site is assumed to be the <u>total</u> deposition of that mineral, the temporal variability need not be concerned in the estimation.

The mean square error (MSE) of an estimator Z_o for Z_0 is defined as $E(Z_o-2_0)^2$• In order to derive MSE, certain knowledge about the covariance structure of the process is required. (Approaches which do not require knowledge of covariance structure are discussed by Creutin and Obled (1982).) First, we describe a few terms.

A 2-dimensional (stochastic) process Z(x,y) is (spatially) <u>weakly stationary</u> if

$$E(Z(x,y)) = m, \text{ and} \tag{3}$$

$$Cov(Z(x,y),Z(u,v)) = K(d) \tag{4}$$

where E(Z) denotes the mathematical expectation of Z, m is a constant independent of (x,y); and Cov(.) is the (spatial) covariance structure depending only on the displacement d between (x,y) and (u,v) through a function K, called <u>generalized covariance</u>. A temporal process (temporally) <u>weakly stationary</u> if

$$E(Z(t)) = m, \text{ and} \tag{5}$$

$$Cov(Z(t),Z(s)) = K^*(t-s) \quad (6)$$

where m is a constant independent of t, and K* is the (temporal) auto covariance function.

This means that the first two moments are invariant under translation. Sometimes the assumption of weak stationary may still be too restrictive, for the hypothesis must be able to cover a broad range of practical situations. Therefore, some 'type of difference' of Z(x,y)are required to be weak stationary. Rodriguez-Iturbe and Mejia (1974)proved that for a process stationary in space and time, the variance of ZQ as defined in (Eq. 1) approaches zero as the number of observations (T) approaches infinity. In other words, ZQ approaches a constant when the time interval between observations becomes sufficiently small.

The assumption of spatial (weak) stationarity can be relaxed to the following:

Def: A spatial process Z(x,y) is said to satisfy the intrinsic hypothesis if Z(x,y) satisfies (Eq. 3) and that its increments are weak stationary:

$$Var(Z(x,y)-Z(u,v)) = 2\ \Gamma(d) \quad (7)$$

where $\Gamma(d)$, as defined, is the (semi-)variogram of Z, and d is the (absolute) displacement between (x,y) and (u,v).

The dependence of Z on d is now through the differences between the observed values, instead of the constant mean. Note that it is not required that the point variance be finite. In practice, $\Gamma(d)$ can be estimated by an empirical variogram

$$\Gamma^*(d) = (1/2N(d)) \Sigma (Z(x_i,y_i)-(Z(u_i,v_i))^2 \quad (8)$$

where the summation is over all N(d) pairs of data points (x_i,y_i) and (u_i,v_i) for which the separation is d. The variogram is usually a monotonically increasing function of d, and is a useful tool in characterizing the process Z (Delhomme, 1978). Sometimes Z is said to be wide stationary if Z satisfies (Eq. 3) and instead of (Eq. 4), the 'correlation' between Z(x,y) and Z(u,v) depends only on the distance d. If the variance of Z(x,y) remains constant at all points, then weak stationary is equivalent to wide stationary.

Matern (1960) reviewed spatial variation of a stationary stochastic process using correlation function. Later, a model for representing the temporal-spatial covariance function of a surface rainfall process, (i.e., rainfall measured at the earth's surface), was proposed by Rodriguez-Iturbe and Mejia (1974), and followed by Mejia and Rodriguez-Iturbe (1974), Bras and Rodriguez-Iturbe (1976a, 1976b, 1976c), Lenton and Rodriguez-Iturbe (1977), Bras and Colon (1978), and Vicens et al. (1979). Recently, Gatz and Naiman (1980) used this approach to investigate the spatial variability of a number

of ions in rain water. The model assumed that the stochastic process is weakly stationary in time and space, and the covariance structure of the process is separable into a temporal component and a spatial component as follows:

$$\mathrm{Cov}[Z(x,y,t),Z(u,v,s)] = \sigma^2 r(d) r^*(t-s) \qquad (9)$$

where σ^2 is the (constant) point variance of $Z(x,y,t)$, d is the distance between (x,y) and (u,v), $r(d)$ is the spatial component of the covariance function, and $r(t-s)$ is the temporal component. In the development, $r^*(t-s)$ has been approximated by a simple Markovian form (Rodriguez-Iturbe and Mejia, 1974), namely,

$$r^*(t-s) = p^{|t-s|} \qquad (10)$$

where p denotes the first-order auto correlation coefficient of the process. In addition, by assuming the process to be isotropic, $r(d)$ can be taken to be either a negative exponential function or a first-order modified Bessel function. (See Bras and Rodriguez-Iturbe, 1976b, for more extensive discussion on the form of $r(d)$.) Rodriguez-Iturbe and Mejia (1974) also presented a geometrical method to fit $r(d)$. From these, estimation of the variance of an estimator for ZQ can be derived.

Estimation of a long-term areal mean. To estimate the areal mean (Eq. 1), Rodriguez-Iturbe and Mejia (1974) used the following simple arithmetic mean

$$Z_0 = (1/NT) \Sigma\Sigma\, Z(x_i,y_i,t) \qquad (11)$$

and they showed that its MSE can be expressed as a product of 3 terms

$$E(Z_0 - Z_0)^2 = \sigma^2 F_1(T) F_2(N) \qquad (12)$$

where σ^2 is the point variance, F_1 is the variance reduction factor due to temporal sampling, and F_2 is the variance reduction factor due to spatial sampling.

The use of (Eq. 11) implicitly assumes that the collectors (stations) were uniformly distributed over the area of interest, each representing area of equal size. This is seldom the case in reality. A better estimator can be defined similarly as in (Eq. 11) by replacing the point depth $Z(x,y,t)$ by areal depth: H, namely, point depth weighted by the size of surrounding area. Rodriguez-Iturbe and Mejia (1974), and Lenton and Rodriguez-Iturbe (1977) further found optimum weights for $Z(x,y,t)$ by minimizing MSE. Those 'optimal' weights are not necessarily the same as the areal depth weights, they can be further improved 'after' $Z(x,y,t)$'s are weighted first by the area size. Bras and Colon (1978) extended the estimation procedure to allow for the covariance function to be non-separable by adopting a stationary (multivariate) autoregressive model for the underlying process.

PERMISSIBLE COVARIANCE AND VARIOGRAM MODELS

According to Christakos (1984), for a p-dimensional function to be a covariance (C), variogram (Y), or k-th order generalized covariance (K) (see below), it has to satisfy some necessary conditions.

1) They belong to the class of real, even and continuous functions (except possibly at the origin).

2) The covariance always has an upper bound: $|e(d)| \leq \varepsilon(0)$. (In reality, the variogram or generalized covariance are rarely so bounded).

3) They behave at infinity according to the laws:

$$\lim \varepsilon(d)/|d|^{(1-p)/2} = 0, \quad \text{as } |d| \to \infty$$

$$\lim \gamma(d)/d^2 = 0, \quad \text{as } |d| \to \infty$$

$$\lim K(d)/d^{2k+2} = 0, \quad \text{as } |d| \to \infty$$

where $d=(d_1,\ldots,d_p)$, and $|d| = (\Sigma d_i^2)^{1/2}$.

In general, for any continuous function to be used as a covariance, (semi)-variogram or generalized covariance model, it is necessary and sufficient that it satisfies the non-negative definiteness property, which ensures the existence of variance of a linear combination of N points,

$$Z_1 = \Sigma \lambda_i Z(x_i) \tag{13}$$

More explicitly, let $h=X_j-x_i$, then

$$\Sigma\Sigma\ \lambda_i\lambda_j C(h) \geq 0, \tag{14}$$

for any n and any complex numbers λ_i; or

$$-\Sigma\Sigma\ \lambda_i\lambda_j \gamma(h) \geq 0 \tag{15}$$

when $\sum Y_i=0$, (the y's form an 'authorized linear combination');

$$\Sigma\Sigma \gamma_i\gamma_j K(d) \geq 0 \tag{16}$$

for all generalized increments of order k (defined below).

A function is called a permissible covariance, variogram or generalized covariance model, if it satisfies (Eq. 14) through (Eq. 16). Permissibility in R^m implies permissibility in R^n for n smaller than m, but the reverse is not necessarily true. In practice, it is difficult to apply these formula to test if a proposed model is non-negative definite and, therefore, a permissible model. The functions fitted to the experimental data are not necessarily among the known permissible models nor models derived from them (see Christakos (1984) for more details).

KRIGING

D. G. Krige used a technique to estimate ore reserves 20 years ago, which later developed into what is now called "kriging" today. Major theory of kriging was formulated and developed by Matheron and others (e. g., Matheron, 1971) at the Ecole des Mines de Paris. The term 'kriging' was introduced by Matheron as an extension of his theory of regionalized variables. The technique, in statistical terms, is an extension of the generalized least squares. Delhomme (1976, 1978) applied the kriging method to hydrosciences, Delfiner (1976), Delhomme (1978), Volpi and Gambolati (1978), Chua and Bras (1981), and Hughes and Lettenmaier (1982) applied kriging to non-stationary spatial processes. Other applications to hydrosciences include Delhomme (1979) to groundwater flow, Villeneuve et al. (1979) to streamflow network design, Volpi and Gambolati (1978), and Karlinger and Skrivan (1981) to estimation of mean annual precipitation over a river basin. General discussion on kriging can be found in Matheron (1971), David (1976), Delhomme (1978), Kafritsas and Bras (1981), and Chirlin and Wood (1982). Recently, Finkelstein and Seilkop (1981) and Finkelstein (1984) applied kriging to a study of acid precipitation.

Basically, kriging is used to find the relationship between the spatial correlation and the precision of interpolation, and/or to estimate the mean of a regionalized variable from a 'single realization' of Z, i.e., no temporal variability is included. If the regionalized variable is estimated at a point, the technique is called punctural kriging. To make statistical inference feasible, certain assumptions about Z must be made. The model assumes that Z is "ergodic", i.e., the first two moments of Z tend asymptotically to certain constants. The variogram $\Gamma(d)$ is allowed to be anisotropic, though the procedure then becomes complicate. In addition, sometimes Z is assumed to satisfy, in order of complexity, spatial (weak) stationarity, the intrinsic hypothesis, or to allow for drift (trend).

Estimation of areal mean. To estimate the regional mean of Z over a region S (Eq. 2), a weighted average of N point observations as defined in (Eq. 13) is used. The estimator is derived so that it is unbiased, and minimizes the MSE. It can be shown that such an estimator exists by solving a 'kriging system[1] of equations (Delhomme, 1978). The estimated weights which satisfy the above conditions are called kriging values, and the minimum variance the kriging variance.

a) Stationary case

In this case, drift is a constant. Assume that Cov(h) is known. The class of estimators have the form Z_1 as defined in (Eq. 13). For this type of linear estimators two conditions are required: unbiasedness and minimum MSE. More explicitly,

$E[Z_1-Z_1]=0$, or equivalently

$$E[\Sigma\lambda_i Z(x_i)] = E[(1/A)\lambda\int Z(x)dx],$$

from which it follows that $EA.=1$, since $EZ(x_i)=EZ(x)$ is a constant.

The MSE is

$$E(Z_1-Z_1)^2 = E[Z_1^2-2Z_1+Z_1^2]$$

$$= \Sigma\Sigma\lambda_i\lambda_j Cov[Z(x_i),Z(x_j)].$$

$$- (2/A(S)) \int_S \lambda_i Cov[Z(x),Z(x_i)]dxdx_i$$

$$+ (1/A(S))^2 \int\int Cov[Z(x_1),Z(x_2)]dx_1dx_2$$

Using Lagrange multipliers, the optimal weights are derived by minimizing

$$F = MSE + 2w(\Sigma\lambda_i-1)$$

with respect to λ_i and w. A kriging system of equations is thus obtained:

$$\Sigma\lambda_j Cov[Z(x_i),Z(x_j)]-w = (1/A(S))\int Cov[Z(x_i),Z(x)]dx, \text{ for all } i, \quad (\Sigma\lambda_j=1),$$

from which it follows that

$$\text{minimum MSE} = (1/A(S))^2 \int\int Cov[Z(x_1),Z(x_2)]dx_1dx_2$$
$$- (1/A(S))\Sigma\int\lambda_i^* Cov[Z(x),Z(x_i)]dx + w^*$$

where λ^* and w* are the solutions of the kriging system. Furthermore, if the variogram is known, then

$$\text{minimum MSE} = - (1/A(S))^2 \int\int\gamma[d(x_1,x_2)]dx_1dx_2$$
$$+ (1/A(S)) \int\Sigma\lambda_i^*\gamma[d(x,x_i)]dx - w^*$$

The first term can be approximated by

$$- (1/N^2) \Sigma\Sigma\gamma^*[d(x_1,x_2)],$$

and the second term by

$$(1/N) \Sigma\Sigma\lambda_i^*\gamma^*[d(x_i,x_j)].$$

where Y* is the empirical semi-variagram.

b) Non-stationary case

In this case, the expectation of Z(x) is allowed to depend on the (p-dimensional) point x through a finite order polynomial, called drift,

$$E(Z(x)) = m(x) = \Sigma a_v f_v(x) \tag{17}$$

in a neighborhood of x, where a_v is unknown but fixed coefficients and the f_v,

v=1, . . . k, are given basic functions, (generally, they are monomials). The drift can be estimated by trend-surface analysis (Watson, 1971). The technique which 'differences[1] the process in order to achieve zero mean (spatial) stationarity is referred to as universal kriging (Delfiner, 1976: Chirlin and Wood, 1982).

Def: Let $Z(x)^a$ be a function on RP. A linear combination of $\sum\lambda_i Z(X_i)$, $x_i=(x_{i1},\cdots,x_{ip})$, is a generalized increment of order k if

$$\sum \lambda_i \left[Z(x_i) + f_v(x_i) \right]$$

$$= \Sigma\lambda_i Z(x_i)$$

for all $v \leq k$

In other words, a generalized increment filters out polynomials of order k or less in the estimation of areal mean, using a particular set of λ's and X_i's. (The definition can be similarly modified for other linear operators.) From the definition, it follows that

$$\Sigma\lambda_i a_v f_v(x_i)=0, \text{ for all } a_v, v\leq k, \text{ and}$$

$$\Sigma\lambda_i f_v(x_i)=0, \tag{18}$$

The variance can be shown to be (Delhomme, 1978)

$$\text{Var}(Z_1) = \Sigma\Sigma\lambda_i\lambda_j K(d(x_i,x_j)) \tag{19}$$

where $d(x_i, X_J)$ is the distance between x_i and x_j, and K is the generalized covariance of order k. Wahba (1980) presented a different algorithm for estimating areal mean using a similar approach.

Suppose that the drift m(x) has a form like (17), and Z_1 (defined in (13)) is a generalized increment of order k. Then

$$\text{MSE} = (1/A(S))^2 \int\int K[d(x_1\ mx_2)]dx_1dx_2$$

$$(2/A(S)) \int\Sigma\lambda_i K[d(x_i,x)]dx$$

$$+ \Sigma\Sigma\lambda_i\lambda_j K[d(x_i,x_j)],$$

where K is the generalized covariance. Using the unbiased condition and Lagrange multiplier (u's), the equations of kriging system became

$$\Sigma\lambda_j K[d(x_i,x_j)] - \Sigma u_v f_v(x_i)$$

$$= (1/A(S)) \int K[d(x_i,x)]dx, \quad i=1,\ldots N$$

$$\Sigma\lambda_i f_v(x_i) - (1/A(S)) \int f_v(x)dx=0,\ v\leq k, \tag{20}$$

from which it follows that

$$\text{minimum MSE} = (1/A(S))^2 \int\int K[d(x_1,x_2)]dx_1dx_2$$
$$- (1/A(S)) \Sigma\int\lambda_i^* K[d(x,x_i)]dx + \Sigma\Sigma u_v^*\lambda_v^* f_v(x)dxu \tag{21}$$

where λ^* and u^* are the solutions of the kriging system. Furthermore, if the variogram is known, then

$$\text{minimum MSE} = - (1/A(S))^2\int\int\gamma[d(x_1,x_2)]dx_1dx_2$$
$$+(1/A(S))^2 \int\Sigma\lambda_i^*\gamma^*[d(x,x_i)]dx + (1/A(S))\Sigma_{ui}^*\int f_v(x)$$
$$- (1/N^2)\ \Sigma\Sigma\Gamma[^*d(x_1,x_2)],$$

and the second term by

$$(1/N)\ \Sigma\Sigma\lambda_i^*\gamma_j^*\ [d(x_i,x_j)],$$

where Γ is the empirical semi-variogram.

During the fitting stages, either variogram has to be calculated or the degree of drift as well as the generalized covariance have to be determined (see Volpi and Gambolati, 1978:, Chua and Bras, 1981). Delhomme (1978) presented a jackknife procedure to check the validity of the fitted model.

A further generalization of the generalized increment is the following:

Def: An intrinsic random function of order k (IRF-k) is a random function whose generalized increments of order k are second order stationary. For an estimator Z_1 (defined in (13)) and its weights satisfying the unbiased conditions (Eq. 21), the random function $\sum\lambda_j Z(X_j+X)$, X in RP, has mean and variance independent of x. Let

$$H(x) = \Sigma a_v f_v(x)$$

where a's are random or constant. Consider

$$\Sigma\lambda_i[Z(x_i+x)+H(x_i+x)]$$
$$= \Sigma\lambda_i Z(x_i+x) + \Sigma\Sigma\lambda_i a_v f_v(x_i+x) \tag{22}$$

Suppose that $f_v(x)$ belongs to a family of polynomials closed under translations, then there exists functions Bs^1 (Delfiner, 1976) such that

$$f_v(x_i+x)= \Sigma B_s^v(x)f^s(x_i),\ \text{where summation is over s that is} \leq k.$$

Then the second term of the right hand side in (22) becomes

$$\Sigma\lambda_i f_v(x_i+x) = \Sigma\lambda_i\Sigma B_s^v(x)f^s(x_i)$$

$$= \Sigma B_s^v(x)\ \Sigma\lambda_i f^s(x_i) = 0$$

using (18). In other words, when working on increments, one is dealing with a whole class of equivalent functions equal to Z(x) up to a polynomial of degree equal to or less than k, and it is this class of equivalence that is called IRF. The kriging system remains the same as above.

APPLICATION TO RAINFALL

Rainfall values from a dense network of 227 raingages near St. Louis area is analyzed using the kriging technique. The values used are 24-hour rainfall totals in the summers (June, July, August) of 971-1975. For each month, an empirical variagram is plotted and an exponential decaying form is fitted to the graph. The estimated parameters of the exponential function are shown in Table 1. β_1 is the intercept, and B_0 is the decaying factor.

Table 1. Parameters of Exponential Function Fitted to Variograms.

Year	Month	d	P_o	β_1
71	June	102.5	.131	.542
	July	92.5	.137	.686
	Aug	97.5	.026	.700
72	June	82.5	.080	.410
	July	57.5	.164	.421
	Aug	67.5	.131	.731
73	June	87.5	.083	.630
	July	57.5	.208	.536
	Aug	62.5	.337	.406
74	June	82.5	.040	.686
	July	47.5	.090	.400
	Aug	117.5	.209	.655
75	June	102.5	.226	.443
	July	102.5	.267	.417
	Aug	107.5	.117	.467

ACKNOWLEDGEMENTS

Part of the research results presented in this paper was supported by the U.S. Geological Survey grant INT 14-08-0001-G-964. The opinions expressed in

this paper did not necessarily reflect USGS's views. The research assistantship provided by the Research Board of the University of Illinois at Champaign-Urbana is gratefully acknowledged.

REFERENCES

Bras, R. L., and I. Rodriguez-Iturbe, 1976a. Network design for the estimation of areal mean of rainfall events. Water Res. Research, 12:1185- 1195

Bras, R. L., and I. Rodriguez-Iturbe, 1976b. Evaluation of mean square error involved in approximating the areal average of a rainfall events by a discrete summation. Water Res. Research, 12:181-184.

Bras, R.L., and I. Rodriguez-Iturbe, 1976c. Rainfall network design for runoff prediction. Water Res. Research, 12:1197-1208.

Bras, R. L., and R. Colon, 1978. Time-averaged areal mean of precipitation: Estimation and network design. Water Res. Research, 14:878-888.

Chirlin, G. R., and E. F. Wood, 1982. On the relationship between kriging and state estimation. Water Res. Research, 18:432-438.

Christakos, G., 1984. On the problem of permissible covariance and variogram models. Water Res. Research, 20:251-265.

Chua, S. H., and R. L. Bras, 1980. Estimation of stationary and non-stationary random fields: kriging in the analysis of orographic precipitation. Report No. 255, Dept. Civil Eng., M.I.T., 171 pp.

Creutin, J. D., and C. Obled, 1982. Objective analyses and mapping techniques for rainfall fields: an objective comparison. Water Res. Research, 18:413-431.

David, M., 1971. Geostatistical ore reserve estimation. Elsevier Scientific Publ. Co., NY, 364 pp.

David, M., 1976. The practice of kriging. Adv. Geostat. in the Mining Industry, M. Guarascio et al., (eds.), , 31-48.

Delfiner, P., 1976. Linear estimation of non-stationary spatial phenomena. Adv. Geostat. in the Mining Industry, M. Guarascio et al.(eds.), 49-68.

Delhomme, J. P., 1976. Applications de la theorie des variables regionalisees dans les scienses de l'eau. Doctoral thesis, Centre d'Informatique Geologique, ENSMP, Paris.

Delhomme, J. P., 1978. Kriging in the hydrosciences. Advances in Water Resources, 1:251-266.

Delhomme, J. P., 1979: Spatial variability and uncertainty in groundwater

flow parameters: A geostatistical approach. Water Res. Research. 15:269-280.

Finkelstein, P. L., and S. K. Seilkop, 1981. Interpolation error and the spatial variability of acid precipitation. Seventh Conf. on Prob. and Statist, in Atmos. Sci., Amer. Meteor. Soc, Monterey, 206-212.

Finkelsten, P. L., 1984. The spatial analysis of acid precipitation data. J. Clim. and Appl. Meteor., 23:52-62.

Gatz, D. F., and D. Q. Naiman, 1980. Spatial variability of rain water impurities in mesoscale events. Syrap., Interm. Range Atmos. Transp. Proc. and Tech. Assess., 1-3 Oct., Gatlinburg, TN, 6 pp.

Hall, A. J., and P. A. Barclay, 1975. Methods of determining areal rainfall from observed data. Predic. in Catchment Hydro., T. G. Chapman and F. X. Dunin (eds.), Australia, 47-57.

Huges, J. P., and D. P. Letternmaier, 1981. Data requirements for kriging: estimation and network design. Water Res. Research, 641-1650.

Kafritsas, J., and R. L. Bras, 1981. The practice of kriging. Report No. 263, Dept. Civil Eng., M.I.T., 107 pp.

Karlinger, M. R., and J. A. Skrivan, 1981. Kriging analysis of mean annual precipitation, Power River Basin, Montana and Wyoming. Water-Resources Investigations 80-50, U.S. Geological Survey, 25 pp.

Lenton, R. L., and I. Rodriguez-Iturbe, 1977. Rainfall network systems analysis: The optimal estimation of total areal storm depth. Water Res. Research, 13:825-836.

Matern, B., 1960. Spatial variation. Statens Skogsfors. Inst., 49,144 pp.

Matheron, G., 1971. The theroy of regionalized variables and its applications. Cahiers du Centre de Morphologie Math., 5, Ecole des Mines, Fantainebleau, France, 211 pp.

Matheron, G., 1973. The intrinsic random functions and their application. Adv. Appl. Prob..

Mejia, J. M., and I. Rodriguez-Iturbe, 1974. On the synthesis of random field sampling from the spectrum: an application to the generation of hydrologic spatial processes. Water Res. Research, 10:705-711.

Rodriguez-Iturbe, I., and J. M. Mejia, 1974. The design of rainfall networks in time and space. Water Res. Research, 10:713-728.

Vicens, G. J., D. S. Grossman, R. R. Evans, S. H. Chua, D. H. Leighton, 1979. Accuracy of mean areal precipitation estimation. Resourse Analysis, CDM Firm, Waltham, MA.

Villeneuve, J.-P., G. Morin, B. Bobee, D. Lebanc, and J.-P. Delhomme, 1979. Kriging in the design of streamflow sampling networks. Water Res. Research, 15:1833-1840.

Volpi, G., and G. Gambolti, 1978. On the use of a main trend for the kriging technique in hydrology. Advances in Water Resources, 1:345-349. 1:345-349.

Wahba, G., 1980. Continuous space series on the plane and in d dimensions. Time Series, 0. D. Andersn, ed., North-Holland Publ. Co., 387-388,

Waston, G. S., 1971. Trend-surface analysis. J. Int. Assn. for Indus. Geology, 3: 215-226.